

## 「訓」情報と「対」頻度・共起頻度から詩句を構造解析し頻出するフレームパターンを抽出する

### 作業手順

- I. 「訓」によるメタデータの作成
  - A. 『学研大漢和字典』の「音訓索引」から訓と字の対を抜き出す。
  - B. 字ごとに訓を整理する。
  - C. 訓を, 日本語の辞書の情報と語形の特徴によって, 品詞に分ける。
- II. 詩句の解析
  - A. 頻出する対を含む詩句を抜き出す。
  - B. 文字ごとに「訓」によるメタデータを付加する。
  - C. 対構造で共起頻度を補強して切る。
  - D. 人手で修正する（作業は常にペアにも反映する）。
  - E. ラベルを持った階層を作る（作業は常にペアにも反映する）。
  - F. 階層構造のデータを蓄積する。
- III. 頻出するフレームを機械的に抽出する。

## I. 「訓」によるメタデータの作成

## A. 『学研大漢和字典』の「音訓索引」から訓と字の対を抜き出す。

とおき		
<sup>13</sup> 遠	1332	←索引はこんな体裁
とおざかる		
<sup>13</sup> 遠	1332	
とおざける		
<sup>13</sup> 遐	1324	
遠	1332	
とおしとする		
<sup>13</sup> 遠	1332	
、 、 、		

## B. 字ごとに訓を整理する。

遠: とおい,とおき,とおざかる,とおざける,とおしとする

## C. 訓を, 日本語の辞書の情報と語形の特徴によって, 品詞に分ける。※多少, 手作業が必要

遠: とおい/Adj      とおき/N      とおざかる/V      とおざける/V      とおしとする/V

※辞書から

※辞書から

※辞書と語形から

※辞書と語形から

※辞書と語形から

## II. 詩句の解析

A. 頻出する対を含む詩句を抜き出す。

水	39	遠	7	波	89	瀾	1	碧
399		87		8		1		43
山	110	高	9	氣	43	象	9	清

B. 文字ごとに「訓」によるメタデータを付加する。

- 別添資料: 「メタデータを付与した対句」

C. 対構造で共起頻度を補強して切る。

- 別添資料: 「共起頻度を対句で補強して切る」

- 一次的な階層構造は機械的にできるかも

[水遠] [波瀾] 碧

D. 人手で修正する（作業は常にペアにも反映する）。

- 対句の片方を修正したら，常にペアも修正するようにする。

[水遠] [波瀾] 碧  
[山高] [気象] 清

E. ラベルを持った階層を作る（作業は常にペアにも反映する）。

[	水	遠	]	[	波	瀾	]	碧
	N	Adj			N	N		
	主	述			N			Adj
					主			述

※ “波瀾”のように部首が共通する漢字2字の並びは複合名詞であることが多い（CHISE-IDSにお伺いを立てればよい?）

	星	41	月	15	動	22	秋	116	山
<b>N (体言)</b>	ほし		つき,げつ				あき,とき		やま
<b>Adj (形容詞)</b>									
<b>V</b>					うごく				
<b>Vt</b>					うごかす				
<b>Adv (副詞)</b>			つきごとに		ややもすれば				

一次的な階層:           [ 秋 山 ]  
                              N    N

二次的な階層: 動    [ 秋 山 ]  
                  Vt       N

動詞           目的語   … ある程度, 機械的にできる?

F. 階層構造のデータを蓄積する。

### III. 頻出するフレームを機械的に抽出する。

#### A. [XY] V/Adj 主 述

- 並列でないXYは、Xが修飾語であることが多いので、[\_Y] に注目。
- ただし、[X上] [X中] などは除く。
- 要するに、主述構造に頻出する□YとV/Adjの組み合わせパターンを抽出するという事。

#### B. Vt [XY] 動 目

- 上と同じ理由で、[\_Y] に注目。
- 動目構造に頻出するVt [\_Y] の組み合わせパターンを抽出する。